CSCI-GA.3033-004

Graphics Processing Units (GPUs): Architecture and Programming

Programming Assignment 2

The serial code was implemented using the backtracking algorithm which took 0.334 seconds to solve a Sudoku puzzle.

I implemented the parallel cuda version following the Genetic algorithm and simulated annealing.

Below are the techniques that I used for improving the performance of the parallel algorithm:

- Used constant memory to store the initial state of the non-empty cells of the puzzle which are not supposed to be re-positioned by the algorithm.
- Used shared memory to access the puzzle within a block which performing computations on the same reducing the global memory access
- The data was copied from device to device in between kernels to save memory bandwidth from un-necessary data copy to host.
- Applied if condition within the kernel where only one thread was supposed to execute the task to avoid global memory access by other threads performing the same computation and writing to the same memory location. (as in kernel init_rsudoku)

The parallel cuda version 1.781 seconds to solve a Sudoku puzzle.

Due high data dependency and complexity this problem is not suitable to be run on GPU, it is better to run it on CPU.